View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Groningen

University of Groningen

A general family of association coefficients

Zegers, Frits Eduard

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1986

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Zegers, F. E. (1986). A general family of association coefficients. s.n.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 12-11-2019

Summary

A general family of binary association coefficients is presented in this study. A family of association coefficients is defined as a class of association coefficients which share some formal properties. Specifically, a family is governed by a central formula basic to each of the members of the family. Specific members of a family are obtained by specific transformations applied to the variables. Such transformations are called quantifications of the variables in this study.

A number of desiderata for association coefficients is discussed first. The desideratum of symmetry requires that the order of the two variables has no influence on the value of the association coefficient. The second desideratum deals with the range on which an association coefficient can take values. It is required that an association coefficient is restricted to a bounded interval. Desiderata with respect to perfect (maximum) association are that the association of a variable with itself is perfect and that perfect association is transitive. Other desiderata are that an association coefficient is zero in the case of statistically independent variables, that, reversely, a zero coefficient implies statistically independent variables, and that an association coefficient is insensitive to transformations of the variables which are admissible in view of the measurement level of the variables and sensitive to nonadmissible transformations. A desideratum with respect to statistical properties of association coefficients is discussed and, finally, it is required that the matrix with association coefficients between a number of variables (the coefficient matrix) is Gramian (positive semi-definite). It is pointed out that many of the desiderata are satisfied if an association coefficient has a Gramian coefficient matrix with identical diagonal entries.

In view of the importance of Gramian coefficient matrices a number of lemmas is presented which can be used to prove the Gramian property of matrices. Two of these lemmas are discussed in detail in Appendix 3.

Some well-known families of association coefficients are discussed. These are the family of Daniels, the family of Cohen, the families of Janson and Vegelius, the family of Zegers and Ten Berge and its chance corrected version proposed by Zegers. A detailed discussion of the

Zegers and Ten Berge family and its chance corrected version is given in Appendices 1 and 3, respectively. Aspects of the Janson and Vegelius families are discussed in detail in Appendix 4. The principles underlying the families are combined in the general family of association coefficients proposed in this study. The central formula of the general family is a simple function of the quantifications of two variables. A quantification is defined as a matrix containing the relevant information of the variable. Three main types of quantifications are distinguished, partly depending on the measurement level of the variables. The central formula can only be applied if the two quantifications are of the same quantification type. If desired, the central formula can be corrected for chance. It is shown that specific coefficients resulting from the central formula or from its chance corrected version all have Gramian coefficient matrices with unit diagonal entries.

Specific quantifications which yield well-known association coefficients are presented. It is shown that the families of Daniels, Cohen, Janson & Vegelius, Zegers & Ten Berge and Zegers are subsets of the general family. Among the best known association coefficients for variables of the same measurement level which are members of the general family are the product-moment correlation, Tucker's congruence coefficient, Spearman's rank order correlation coefficient rho, Kendall's rank order correlation coefficient tau and Cohen's Kappa coefficient. Special consideration is given to association coefficients for binary (or dichotomous) variables. Well-known examples are the phi coefficient and the G coefficient of agreement. Among the members of the general family for variables of different measurement level are the SP and the CP coefficients of Janson and Vegelius.

8100 dsc 004S 001
8100 208 7292
8100 65286/1988
7800 069979111
7003 11-11-04 : zB
7100 488 9336
8100 04037986
8200 258858354
7800 429950476